DOCUMENT RESUME

ED 047 013 TM 000 388

AUTHOR	Byrd, Manford, Jr.
TITLE	Testing Under Fire: Chicago's Problem.
INSTITUTION	Chicago Board of Education, Ill.
PUB DATE	Oct 70
NOTE	15p.; Paper presented at the Conference of the Educational Records Bureau, "Testing in Turmoil: A Conference on Problems and Issues in Educational Measurement," New York, New York, October 1970
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*City Wide Programs, *Educational Accountability, Norms, Racial Factors, *Standardized Tests, *Test Construction, *Testing, Testing Problems, Testing Programs, Test Interpretation, Test Selection
IDENTIFIERS	*Chicago, Illinois

ABSTRACT

The history and development of city-wide testing programs in Chicago since 1936 are reviewed and placed in context with the impact on testing of Sputnik and the passage of the National Defense Education Act of 1958. Current testing problems include the time lag between events and curricular changes and new test construction, the time lag between changes in student population and re-norming of tests, comparability of school populations of large cities and norms groups, reinforcement by standardized tests of the concept of a standard as the goal rather than progress, and the failure on the part of the public, despite all efforts to the contrary, to appreciate the importance of the standard error of measurement. Among other issues considered are the problems of cultural differences, the results of deprivation, language difficulties, the use of test results as one method of evaluation in accountability, and calls for a moratorium in testing. Suggested remedies include the construction of better, more relevant tests; more frequent test revisions; more sophisticated understanding, interpretation, and use of test results; periodic review of material entered into a student's cumulative record; moderation of test publishers' oversell; and working together to put standardized testing back into context. (DG)

ED047013

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Testing Under Fire: Chicago's Problem

Introduction

Members of this conference:

I am pleased to be here and to try to make a contribution to the deliberations going on during these two days, over coffee, over lunch, and in the work sessions. I feel myself to be a spokesman of the Chicago public school system which tests some 250,000 pupils in city-wide surveys annually. This figure does not include many varieties of testing in ESEA programs, classroom testing initiated by teachers, or the norming that test publishers in increasing numbers are requesting us to do. As an extensive user of tests, both in administration of the instruments and in utilization of the results, we have a major stake in the future of standardized tests and testing. I might add that in this position we are among those on the edge if not in the center of the turmoil.

When I first read the theme of this conference, "Testing in Turmoil: A Conference on Problems and Issues in Educational Measurement," I said to myself, "Why shouldn't testing be in turmoil? Everything else connected with education is!" It is significant, I think, that the turmoil about testing centers about the elementary and secondary education establishment. As far as I know, so far there has been no massive attack upon testing in business and industry; and as my colleagues who are bending every effort to increase minority group graduates to enter higher education know, there is relatively little challenge to the use of cutoff scores for college admissions purposes.

Chicago's Experience

The Chicago public schools instituted a city-wide testing program in the elementary school grades in 1936 and in the high school grades in 1937. From

TM 000 388

the inception of the standardized testing programs until about 10 years ago, each school chose its tests from our official list. To prepare this approved list our Bureau of Instruction Materials first notifies all test publishers of the review. They submit samples and if requested appear before the review committee to explain or defend their materials. The committee is composed of teachers, counselors, and administrators. The list of tests recommended by the committee is then presented to and formally approved by the Board of Education. The list is always very extensive and is used by principals to select tests for whatever testing they desire in addition to the city-wide survey testing.

With implementation of the National Defense Education Act, we began both electronic processing of output and a city-wide adoption of the tests used, with the selection being made from the approved list by a comparable committee process. This change took about seven years to complete. We test at six points, first, third, sixth, eighth, ninth, and eleventh grades. Actually in this conversion we worked downwards, beginning with the high school grades, the costs of which were originally borne by NDEA. In the 34 years of our standardized testing program, there have been some other changes such as the grades tested, instruments or editions used, et cetera. When the sixth and eighth grade programs were to be converted in 1961, and a city-wide committee reviewed the tests on the approved list and recommended selections, I am told the majority opinion came under fire. I mention this only to indicate that way back then when things were relatively quiet, there was turmoil in testing. The program was reviewed a few years later and is up for review again now. The current review committee will be expected to deal with issues and problems and policy recommendations which previous committees felt no need to pursue.

Beginning around 1962 or 1963, pressure began to be put upon the school administration to make test results public. As a result, presentations were made at Board of Education meetings, which are public meetings of course, with the press in attendance. City medians alone were given at first; then information about administrative districts coded for anonymity. When the request came repeatedly for information on a school by school basis, information was presented on five schools, again coded for anonymity, but carefully and methodically selected to represent median school performance from high to low. Some other presentations were tried. Each was very carefully worked up with slides and a simple text. The text was not always distributed, but it was always prepared with as much attention to explicit explanation and grace of style as possible. Last year for the first time, we issued median test scores school-by-school for each grade level tested during the previous school year. Each of our 500 and some schools had a page to itself, and as a result the book is known to some as "the telephone book." To others it is the "green dragon" or the "green monster" in deference to the color of the cover but in tune with some attitudes toward the release.

The book has a substantial introduction which, if read, cannot help but put test results into context - where they belong but do not always find themselves. One final point about that introduction and about the previous more limited presentations of data: the staff responsible tried to make the text a teaching instrument. Those in charge see themselves as teachers still, though the classroom for them is a thing of the past, and they do not boggle at trying to teach Board members, reporters, or the general public which is, after all, usually dependent on the reporters. The results were discouraging in the extreme at the start, but the staff never gave up hope or the teaching role, and in the last few years some light has been seen

at the end of the tunnel. It does appear that the fact is getting through that half of the norm group was below the median or in the frequently misunderstood vernacular, the grade level equivalent. This year our city-wide school-by-school report will show data for two years - a mighty challenge to those who know there will be grumbling if the pages are crowded with figures - as they will be!

I have taken this much time for this overview for a few basic reasons. You need to know the experience of the school system from which I speak (although I wasn't there to live through all of it!). I think this review pretty well summarizes the experience of other school systems, some of which have had far more pressure put upon them than we have. The review carries us from the days when turmoil for testing was nonexistent; when it began as an in-house affair over which instrument to use; when steps were necessarily taken as pressure to release data began; to point where we are now - almost. I say "almost" because a new pressure is developing, not from the black community at present but from the Puerto Rican. But more on that in another context later.

NDEA and Sputnik

I think it is also important to cite the role in the development of school standardized testing played by the National Defense Education Act of 1958. There is no question as to its impetus for testing. I understand that its redundant sub-title, guidance, counseling and testing was necessary for legislators to be satisfied that testing would be included. While aid to vocational education antidates the National Defense Education Act by several decades, NDEA is usually credited with being the forerunner of the massive categorical aid to education during the 60's. It is ironic that this important legislation passed both houses in the name of national defense. NDEA was, I believe,

the first positive response of the federal administration to Russia's Sputnik that lofted in October 1957. There had been previous negative responses when the federal administration castigated the public high schools of the country because we did not get into orbit first.

In this connection, a passage from Arnold Toynbee's book of 1948, Civilization on Trial, seems very pertinent. I think we must all today in our trouble take care it does not apply to us. The passage begins thus: "It is always a test of character to be baffled and 'up against it,' but the test is particularly severe when the adversity comes suddenly at the noon of a halcyon day which one has fatuously expected to endure to eternity." I skip a sentence now and continue: "Yet 'to pass the buck' in adversity is still more dangerous than to persuade oneself that prosperity is everlasting." That is the end of the quotation, but it is not the end of the idea.

The buck was passed to public education, especially secondary education, following Sputnik, and adversity has endured since then. Despite the fact that three months before Sputnik, the Board of Education in Chicago adopted a much more stringent high school graduation requirement after 12 to 18 months of study has not relieved it of adversity of any sort. The causes of adversity have, in fact, compounded since then. For the last several years, the results of school standardized testing have brought the schools under fire. With every major city in the country in trouble when its results are compared with the norm group, one begins to ask where the problem does lie. What is its nature? What are its dimensions? I want to turn now to some of these problems and issues without presuming to exhaust either roster. In fact, in organizing these remarks, I found it difficult to be sure always which point is a problem and which is an issue.

Problems

I will start off with what is a difficult problem but one which I think is only by inference an issue. That is the problem of lag and of the inability of so ponderous an enterprise as test construction to keep abreast of the changes and realities. Let me begin with an illustration from World War II days and individual psychological testing. I am told that the Wechsler of that day had as an easy question the query as to the name of the previous president of the United States and as a difficult question, phrased in some form, "Where is Tokyo?"

In the 1940's after war began with Japan, these questions virtually changed places on a scale of difficulty. Every school child with a brother, uncle, or father in service knew where Tokyo was. Conversely, with Roosevelt in the presidency since their infancy or early childhood, Hoover was a name unknown to them. You will get my point but let me make two more that represent lag.

One has to do with the new math, which of course is new only where the old math is no longer being taught. Now, on this one, I am myself confused. It seems to me that whether the method is new or old, the purpose is always to find the answer and that either method should lead to the resolution of the purpose. To believe otherwise is as much as to say that a problem must be stated in a single form to be soluble. Nevertheless, there is a problem for test builders in keeping abreast of curricular changes and of satisfying the needs of all school systems across the country.

Those of us in the business know that test builders study the curriculum guides across the country and the textbooks too. We understand the built-in lag. It has taken several years, for example, for textbooks to carry illustra-

tions of non-white faces to say nothing of introducing appropriate mention of ethnic contributions to our national life. We know that test changes follow curriculum and textbooks changes,but our vocal friends do not.

The dozen or so major cities where most of the confrontations take place buy a great many tests, but they do not dominate a company's sales. Big city constituencies just do not understand this.

I want to cite one final and perhaps more devastating problem from this generalized problem I have identified as lag. The problem may very well be confined to the major cities, but I do not think so. Suburbs are also now experiencing it, and that is the matter of the changing population.

Leona Tyler says in her book of 1965, entitled The Psychology of Human Differences, "We know now that test scores can be manipulated to give us any sort of distribution we want. Because there are definite mathematical advantages to be obtained from normal distributions, one of the aims of present-day test-builders is the construction of tests that will give normal distributions for the types of population in which they are to be used." Let me repeat: "Normal distributions for types of population in which they are to be used."

A question being raised today which brings the testing program under fire is quite simply,"Are the school populations of large cities comparable to the norm groups?" That is to say there is the feeling that with the changing population in great cities the old norms are not relevant. It is really difficult to conclude otherwise when every major city is in trouble with its constituency because its median falls short of the national norm. Enlightened believers in testing are beginning to question not testing or its value but the tests, but not so the public. It questions the quality

of the educational program. This problem is compounded by the fact that the public still believes the median represents what every Tom, Dick and Harriet not only should but must achieve for a school to be doing its job.

I am afraid that, unwittingly, standardized testing has reinforced the concept of a standard as the goal rather than progress as a goal. Now, you and I know that statistically everyone cannot score at the median, which in achievement tests is most often, however erroneously, translated into grade level equivalents. I once heard a well intentioned, but misguided administrator declaim that every pupil had to achieve his grade level equivalent, not realizing he was talking of a statistical derivation of a median. As I listened, I thought I heard that median bounce up against the ceiling with the staccato echo of a ping pong ball. I did, in fact, glance up to see if there was a dust mark on the acoustical tile overhead. I understand that the distinguished consultant of one test company with a revision in the works is advocating dropping it, and I hope he prevails and wins converts in other companies. It is tiresome to keep repeating the truth about that shibboleth including the way performance it seems to denote actually varies from subtest to subtest.

I just do not know how all of us working together are going to put it across that a child's progress is more important than his standard of achievement. Take, for example, a sixth grade pupil who tests at the fifth grade level and who in eighth grade tests at the seventh grade level. He still is not up to the artificial standard, but he has made steady progress which is lost to sight because of/imposition of a standard as a measure of the achievement rather than progress. Of course we compound this problem when we publish school-by-school test data that obscure the individual achievement of pupils inasmuch as the public seems to rivet its attention on the

bottom half almost forgetting the top half.

Closely related to this point is the problem posed by the American confidence in a number. Despite our efforts in Chicago to enlighten our public about test scores, I doubt we have succeeded in shaking belief in the sanctity of the numbers and in the infallibility of a test. When we try to explain that one has to bear in mind the standard error of measurement, the idea is brushed aside, perhaps because the public has to believe in the certainty of a number and cannot tolerate the slightest slippage of confidence. We provide stanines and percentile ranks as well as grade equivalents but in our public school-by-school reports we do not include stanines as being too gross for large groups. They are apparently readily understood by parents and pupils and in them we adjust for the standard error measurement without having to identify it.

There are some problems of format, quality of printing of tests, and the sizes of the targets that must be blackened. These may seem to be minor problems and in many ways they are. They do seem as if they are correctable, however.

Other problems center around the cultural differences among children, the results of deprivation, and language difficulties, but these have become issues, and now I turn to them.

Issues

The problems of cultural differences, the results of deprivation, vocabulary, and language difficulties first came to the fore with respect to our black population. Everyone here is familiar with the efforts to develop culture free or culture fair tests and the apparent impossibility of altering results. In my own experience when my school participated in

the norming,no group's relative position was changed as a result. A few years ago Chicago ran an experiment with culture fair tests only to discover that the deprived children did better on the test in the city-wide survey than on the experimental instrument. It is a dilemma for us all, but the response of the black community was to grumble about the issue and to put the pressure on the school for better teaching.

Now the pressure is coming from the Mexican-Americans and the Puerto Ricans and has been for some time. Of these voices, the Puerto Ricans' dominate. They raise the issue not only of cultural impositions but also of our expectations as to competence in English,and they are firing at the testing program, which they say penalizes Puerto Rican children because of cultural differences and problems of English. We have found, of course, that many Puerto Rican children who cannot read English cannot read Spanish either.

All of this has become an issue of civil rights, and a Health Education and Welfare directive has in fact been issued concerning placement of minority origin children in special classes on the basis of examination where a knowledge of English is crucial and concerning access to college preparatory courses. The implications for testing are enormous, and our Spanish speaking fellow citizens prefer not to recognize the fact that textbooks take time to produce as do tests which are contingent upon national curricula and content. Perhaps in this connection it would be interesting to you to point out two cultural differences they cite. Puerto Ricans claim test situations put a premium on the ability of their children to work effectively within time limits which is part of the culture of continental Americans but not of Puerto Ricans. They say continental Americans are future-oriented but Puerto Ricans are not. In

their pressure for alterations in testing and test use, they are demanding we reverse their "now" orientation and our future orientation thus creating a paradox on the one hand and a dilemma on the other; mañana for them but today for us. Our Puerto Rican Board member has begun to vote against test contracts in general to emphasize her point of view.

Until recently no voices were raised in questioning one goal of education - to transmit, even to help create a common culture, to make us one. Today diversity is in. While once the perpetuation of distinctive ethnic or national traditions and history was left to the home, the school is being expected to do the job today. As one point of interest in this regard, the Illinois School Code was amended in June 1967 to add this passage in the section (27-21) pertaining to the teaching of United States history:

"The teaching of history shall include a study of the role and contributions of American Negroes and other ethnic groups including but not restricted to Polish, Lithuanian, German, Hungarian, Irish, Bohemian, Russian, Albanian, Italian, Czechoslovakian, French, Scots, et cetera, in the history of this country and this State."

As I read over that passage in preparing this paper, I was struck by the fact that the American Indian was again consigned to et cetera in history and is still dependent upon the inference drawn by the sensitive reader. Chicago has a sizeable Indian population as off-reservation populations go, and the voice of the American Indian is the latest to be raised. The objections and issues voiced by them pertain first and mainly to the curriculum and to teaching expectations and methods, but they all have implications for testing even when not stated.

These voiced dissatisfactions pose grave problems to educators but they are test publishers' problem also. Neither of us can quiet the issues alone. We cannot pass the buck between us but the halcyon days have departed. We need each other to work through our mutually critical situation.

Another issue pertains to the use of test results as one method of evaluation in accountability, the latest intrusion of the business and industrial world into education. Now, of course, every teacher and administrator has to be accountable. He must be professional, diligent, creative, and sensitive, but he is dealing with human beings who come from environments and backgrounds over which he has no control. For several years recently, the concept of "Zero Reject" caught the fancy of some. Of course, the school should not reject anyone, but the school is not a production line where the input as well as the output can be controlled, as well as all the influences in between the two. Education, if it is worth anything at all, must be multi-dimensional. Reading, writing, and arithmetic competence are terribly important but they do not constitute an education or the totality of growth. Many of the other elements of an education are unmeasurable at least today; moreover, they come to true fruition long after a youngster has left school. The public school educational effort in this country should not be condemned; what it needs is to be assisted. It needs to be assisted not only by truly massive financial aid to education but also by attention to housing, to open occupancy that is a fact, to truly fair employment practices, and methods of upward mobility.

What do these have to do with testing? Everything. Testing is like the tail of the dog. Tails droop and wag. They also sometimes get lopped off.

It would surely be a setback if standardized testing in schools were to be discontinued. I remind you of what has happened in some places. Chicago has not as yet been fully under this threat, but some voices are suggesting the threat. It is worth mentioning that at the 1970 spring American Personnel and Guidance Association convention in New Orleans a resolution was adopted calling for a moratorium on testing. The resolution was not specific to schools but referred to "certain individuals and organizations" who use tests "to limit the job opportunities and educational development of low socioeconomic, underprivileged, and non-white individuals." It outlined some steps for improvement in understandings, materials and research techniques and ended by resolving that if "demonstrable progress in clarifying and rectifying this situation cannot be achieved by this time next year, proposals for a moratorium on use of group intelligence tests with these groups be presented." Schools would, I believe, regret the necessity of this drastic antidote. There is no assurance that the medicine would not be more toxic than the infection, for one thing. For another, achievement tests are as capable of labeling as are so-called intelligence tests.

Remedies

I touch very sketchily upon remedies, for the fire must be extinguished and the turmoil quieted. First there is no question that we need better tests - better in content and format. They must be made more evidently relevant and somehow they must be more quickly responsive. I do not know how the last two can be achieved. Nor if they are achieved, do I know how school systems like Chicago's, always on the verge of impoverishment, can in turn respond with the repeated purchase of new booklets, to strike a very practical note. Changing tests frequently has other obvious disadvantages. I guess I am saying that

one solution is to find a way to do the nearly impossible but we know doing that merely takes a bit more time - and frequently, more money.

Second, we need more sophisticated understanding, interpretation and use of test results by teachers, counselors, and administrators. Somehow we all, working together, have to decanonize those numbers and return to some confidence in our professional assessments, our professional judgment. In a sense there has been an abdication in favor of numerals as dictator, a handy fellow to whom to pass the buck of responsibility.

We also have to review and decide anew periodically what goes into a pupil's cumulative record that follows him like a shadow. In this regard, the recent splendid report of the Russell Sage Foundation, entitled, Guidelines for the Collection, Maintenance and Dissemination of Pupil Records should be required reading for us all - with some results thereafter.

Finally, all of us together have to put standardized testing back into context. For example, a 1968 handbook called Guidance Services for Illinois Schools includes a section headed "Guidelines for Developing a Testing Program" which states "Keep in mind: A test is but an indicator of a pupil's performance on a given day and under a given set of circumstances." Most of that quotation is printed in bold type. There are other remedies, but I doubt if there are more categories for them: better instruments, more sophisticated use, and wider perspectives on testing.

Conclusion

I began these remarks by facetiously suggesting that standardized testing, as part of the educational establishment, had its turn in turmoil coming. In concluding, I return to that point without any intended humor. Just as the school is indirectly being held accountable for the results of social deficiencies in this country, it is also being trapped, directly, by a sort of

overkill, a kind of oversell in testing. I do not want to see a moratorium on testing - in schools at least. I do not want to throw the baby out with the bath, to employ a repulsively trite but very graphic cliché.

What I would like to see is a return to moderation on the one hand and to responsibility on the other. Test publishers have to moderate their oversell. Test users have to upgrade their insights and return to their responsibilities. Test publishers have to assume more responsibility for ensuring adequate interpretation and appropriate use of test results, with for example, better descriptions of the norm group. Test users must moderate their reliance upon test results.

In short, we have to douse the fire and quiet the turmoil by some united professional approach. Testing is lucrative big business. Education is big business, but it is not financially lucrative. Education is, however, not only the highway to the gross national product and dividends of separate companies but also the gateway to the American ethic.

Neither testing nor education is isolated in turmoil today, in this country or elsewhere. Newspapers, the radio, and television never let any of us forget that fact. Thus we cannot escape the crisis of fire or turmoil. We can, however, work together to overcome our deficiencies and to bulwark our strengths. Indeed we must; but we have to have the sense to discern the difference between them and the integrity to act on our own collective discernment.

Such is Chicago's problem which is also yours. Such is Chicago's goal which I hope you share.